

Optimization Performance of Extreme Gradient Boosting and Random Forest for Child Stunting Classification Based on Economic Factors

Yuyun Yusnida Lase*¹, Purwa Hasan Putra², Arif Ridho Lubis³, Santi Prayudani⁴

^{1,2,3,4} Teknologi Rekayasa Perangkat Lunak, Politeknik Negeri Medan, Indonesia

Email: yuyunlase@polmed.ac.id

Received : Apr 16, 2026; Revised : May 6, 2026; Accepted : May 7, 2026; Published : Jun 15, 2026

Abstract

Stunting remains a major health concern in Indonesia due to its impact on children's physical growth and cognitive development. One of the factors influencing the incidence of stunting is family economic status, which is linked to access to nutrition, sanitation, and a healthy environment. This study aims to optimize the performance of the XGBoost and Random Forest algorithms in classifying stunting in children based on economic factors and to compare the performance of the two models. The methods used in this study involve a machine learning approach, including data preprocessing, model training, hyperparameter optimization, and performance evaluation using a confusion matrix, accuracy, precision, recall, F1-score, and ROC-AUC curves. The results indicate that both algorithms perform well in classification, with an accuracy rate of approximately 70%. The Random Forest model demonstrated better performance than XGBoost with an AUC value of 0.7655, while XGBoost had an AUC value of 0.75. Additionally, the feature importance results indicated that economic and environmental factors, such as housing conditions and sanitation, have a significant influence on the incidence of stunting.

Keywords : *Child Stunting, Economic Factors, Random Forest, XGBoost, Optimization*

This work is an open access article licensed under a Creative Commons Attribution 4.0 International License.



1. INTRODUCTION

Stunting is a chronic nutritional problem that remains a global and national challenge, particularly in efforts to achieve the Sustainable Development Goals (SDGs) [1]. Stunting remains one of the most serious chronic nutritional problems affecting children, as it impacts physical growth, cognitive development, productivity in adulthood, and the quality of human resources. Although various interventions have been implemented, the prevalence of stunting in many developing countries, including Indonesia, remains high. This situation indicates that stunting is not merely a health issue but is also closely linked to complex social and economic factors [2], [3], [4].

One of the main factors contributing to stunting is a family's economic situation. Low economic status can limit access to nutritious food, health care, sanitation, and adequate education [5], [6]. This situation means that children from low-income families are at a higher risk of stunting than children from families with better economic conditions. Therefore, economic factors are an important variable that must be taken into account in the analysis and classification of stunting risk [7], [8].

Household economic status is one of the primary determinants of stunting in children. Economic constraints lead to limited access to nutritious food, health care, education, and adequate sanitation. Recent research indicates that children from families with low economic status face a higher risk of stunting compared to children from more affluent families. Furthermore, economic factors interact with other variables such as parental education, environmental conditions, and access to healthcare facilities, creating a complex pattern of relationships that is difficult to analyze using conventional statistical methods [9], [10].

As technology advances, machine learning approaches are increasingly being used in the healthcare sector to model and predict the risk of stunting. These methods are capable of capturing nonlinear relationships and complex interactions between variables that traditional methods cannot adequately account for. Recent studies indicate that ensemble-based algorithms such as Random Forest and XGBoost demonstrate superior performance in classifying and predicting stunting due to their ability to handle large datasets, diverse features, and reduce overfitting [11], [12].

The urgency of this research lies in the importance of developing an accurate and efficient classification model to support decision-making in efforts to prevent stunting. With the increasing volume of health and socioeconomic data, there is a need for analytical methods capable of processing this data optimally so that it can provide relevant information to policymakers [13], [14].

To address this issue, this study proposes an approach to optimize the performance of the XGBoost and Random Forest algorithms in classifying child stunting based on economic conditions. The optimization process involves parameter tuning to improve the accuracy and performance of the models. Additionally, a comparison between the two algorithms is conducted to determine the most effective method in the context of the data used [15], [16].

Several previous studies have examined the application of machine learning methods in the classification and prediction of stunting in children using various algorithmic approaches. This study developed a stunting classification model using three algorithms: Naïve Bayes, K-Nearest Neighbors (KNN), and Random Forest [17], [18]. The dataset used consisted of stunting data from children in Indonesia, comprising 10,000 records that included variables such as age, weight, height, and breastfeeding history. The results of the study indicate that the Random Forest algorithm delivers the best performance, with an accuracy rate of 87.75% and an F1-score of 0.922. This demonstrates that Random Forest is capable of providing a good balance between precision and recall in detecting cases of stunting [19], [20].

Research using Random Forest on stunting data in East Java achieved an accuracy of up to 93%, while the Naïve Bayes method achieved an accuracy of 87.3%. Meanwhile, the KNN approach with parameter optimization achieved an accuracy of up to 92.20% in classifying child stunting in Indonesia. This indicates that decision tree-based and ensemble algorithms perform quite well in addressing stunting classification problems [21].

This study focuses on the application and optimization of machine learning algorithms for classifying stunting risk, with an emphasis on economic factors as the primary variables. Through model optimization, it is hoped that a more accurate classification system can be developed, thereby assisting healthcare workers and policymakers in early detection and in determining more targeted interventions to address stunting.

2. METHOD

This study employs a quantitative approach using experimental methods in the field of machine learning to optimize the performance of the XGBoost and Random Forest algorithms in classifying child stunting based on economic factors. The research stages were conducted systematically, from data collection to model evaluation.

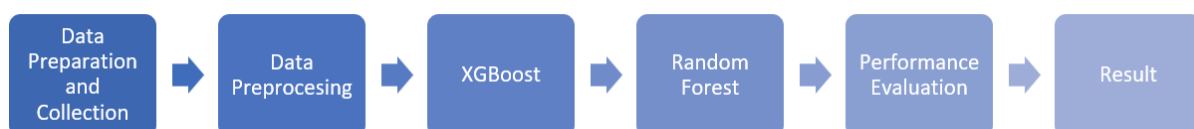


Figure 1. Research Flow

2.1. Data Preparation and Collection

In this study, data collection was conducted by obtaining secondary data from a government agency, namely the Department of Health. The data collected pertained to children’s nutritional status, particularly stunting, as well as socioeconomic information about families, which was the focus of the study.

The variables used included economic indicators such as family income, parents’ educational level, and occupation, as well as other supporting variables relevant to children’s health.

2.2. Data Preprocessing

Before modeling begins, the data first undergoes a preprocessing process that includes:

Data cleaning to address missing values and duplicate data

Data transformations such as encoding for categorical variables

Data normalization or standardization, if necessary

Splitting the data into training and testing sets, typically in an 80:20 ratio. This step aims to ensure data quality so that the resulting model is more accurate [22].

2.3. Optimization of the XGBoost and Random Forest Models

This study uses two machine learning algorithms, namely:

XGBoost (Extreme Gradient Boosting): a boosting algorithm that combines multiple weak models to improve predictive performance

Random Forest: an ensemble learning-based algorithm that uses multiple decision trees to improve accuracy and reduce overfitting Both models were trained using training data to identify patterns in the relationship between economic factors and stunting [23].

2.4. Performance Evaluation

Evaluation metrics are created in a similar manner to the process followed in binary and multiclass classification, specifically for your Label task. These metrics provide a quantitative assessment of the model's performance in handling multiple classes, providing insight into aspects such as precision, recall, F1 score, and accuracy for each class in binary and multiclass classification problems. This provides a solid basis for evaluating the effectiveness of the model in distinguishing between different classes in multiclass and binary classification databases [24], [25].

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Prec = \frac{TP}{TP+FP} \quad (2)$$

$$ec = TP_{rate} = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - Score = \frac{2}{precision^{-1}+recall^{-1}} = 2 \times \frac{precision \times recall}{precision+recall} \quad (4)$$

3. RESULT

This section covers the research process from the initial data collection, through data preprocessing, to the implementation of the XGBoost and Random Forest optimization models. Performance evaluation will then be conducted using ROC, accuracy, precision, recall, and F1-score.

3.1. Dataset

The data collection process for stunting in this study began with coordination with the Health Office, which served as the primary data source. The data obtained consisted of secondary data derived from routine records at health care facilities such as community health centers and integrated health

posts. At this stage, the researchers submitted a request for access to data regarding children’s nutritional status and family socioeconomic information.

Table 1. Dataset Stunting

No.	Age	BB	TB	Status Home	Condition of the House	Sanitation	Drinking water facilities	Stunting	
0	1	16	6,8	68	Oneself	Unhealthy	Unacceptable	Bottled water	1
1	2	52	11,6	95	As a guset	Unhealthy	Unacceptable	PDAM	1
2	3	32	10,5	85	Rent	Unhealthy	Unacceptable	PDAM	1
3	4	52	11,6	95	As a guset	Unhealthy	Unacceptable	PDAM	1
4	5	32	10,5	85	Rent	Unhealthy	Unacceptable	PDAM	1

Next, data is collected through direct measurements conducted by health workers in the field. Measurements included the child’s age, weight, and height, which were then used to determine stunting status based on WHO standards. In addition, staff also collected data on socioeconomic conditions, such as housing status (owned, rented, or living with others), housing conditions (healthy or unhealthy), sanitation, and access to drinking water.

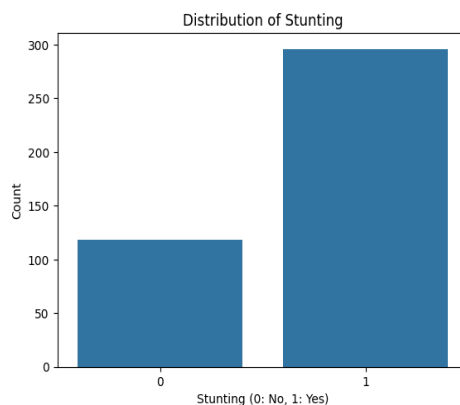


Figure 2. Data Distribution

3.2. Data Preprocessing

Table 2. Results of Stunting Dataset Preprocessing

No.	Age	BB	TB	Status Home	Condition of the House	Sanitation	Drinking water facilities	Stunting
0	1	16	6,8	68	3	0	1	1
1	2	52	11,6	95	1	0	3	1
2	3	32	10,5	85	4	0	3	1
3	4	52	11,6	95	1	0	3	1
4	5	32	10,5	85	4	0	3	1

The data preprocessing stage is a crucial step in this study, aimed at preparing the data for use in machine learning modeling. The initial step involves data cleaning, which entails removing missing values, duplicate data, and input errors. For example, if there are missing or illogical values for weight or height, appropriate actions—such as data deletion or imputation of specific values—are taken to prevent these from interfering with the analysis process.

Table 3. Results of Stunting Dataset Preprocessing

	No.	Age	Status Home	Condition of the House	Sanitation	Drinking water facilities	Stunting
count	592.000000	592.000000	435.0	565.0	569.0	592.0	592.000000
mean	296.500000	33.788851	2.944828	0.679646	0.72232	2.295608	0.775338
std	171.039956	13.869259	1.22632	0.467026	0.448249	0.811579	0.417713
min	1.000000	1.000000	1.0	0.0	0.0	1.0	0.000000
25%	148.750000	23.000000	2.0	0.0	0.0	2.0	1.000000
50%	296.500000	34.000000	3.0	1.0	1.0	3.0	1.000000
75%	444.250000	45.000000	4.0	1.0	1.0	3.0	1.000000
max	592.000000	62.000000	4.0	1.0	1.0	3.0	1.000000

3.3. Optimization of XGBoost and Random Forest

The results of the study indicate that stunting data obtained from the Health Department can be effectively used to build classification models using the XGBoost and Random Forest algorithms. After undergoing preprocessing steps, such as data cleaning, handling missing values, and variable transformation, the dataset became better prepared for analysis. The model training process indicates that both algorithms are capable of identifying patterns of relationships between family economic factors and children’s stunting status. This suggests that economic variables such as income, parental education, and employment play a significant role in determining the risk of stunting.

3.3.1 XGBoost

The results of the XGBoost optimization indicate that the parameter tuning successfully improved the model’s ability to recognize data patterns, particularly in detecting cases of stunting. This aligns with XGBoost’s characteristic of being able to iteratively improve prediction accuracy through boosting techniques.

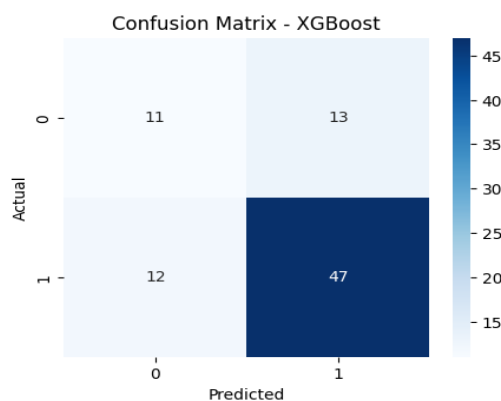


Figure 3. XGBoost Confusion Matrix Results

The results of the XGBoost model implementation show that the model is capable of classifying stunting status with fairly good performance. Based on the confusion matrix, the True Negative (TN) value is 11, the False Positive (FP) value is 13, the False Negative (FN) value is 12, and the True Positive (TP) value is 47. From these results, it can be seen that the model is more effective at classifying children with stunting than those without stunting, as indicated by the high True Positive value. Overall, the model has an accuracy rate of approximately 70%, indicating that the majority of the data was classified correctly.

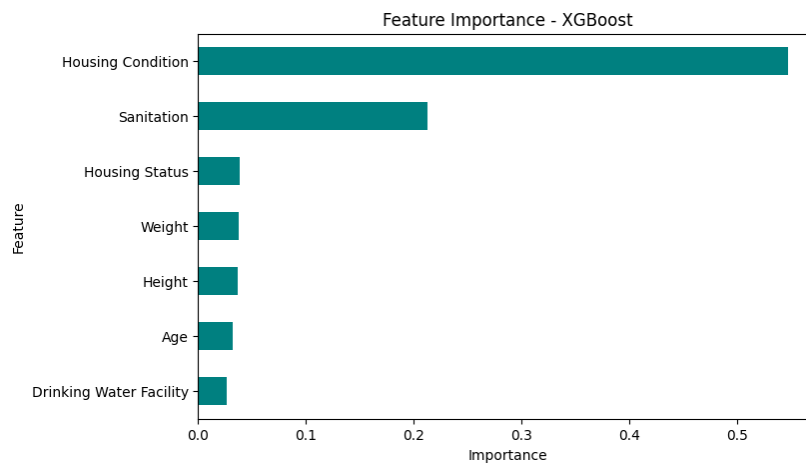


Figure 4. XGBoost Feature Importance Results

The feature importance results show that the XGBoost model is capable of identifying the most influential variables in the classification of stunting based on economic factors. According to the graph, housing conditions are the most dominant factor, with the largest contribution compared to other variables. This indicates that the quality of the living environment—such as housing density, ventilation, and housing adequacy—has a highly significant impact on the risk of stunting in children. Furthermore, sanitation variables also have a considerable contribution, indicating that access to good sanitation plays a crucial role in preventing growth impairments caused by infections or an unhealthy environment.

3.3.2 Random Forest

The results of implementing the Random Forest model demonstrate fairly good classification performance in identifying stunting status in children. Based on the confusion matrix, the values obtained are True Negative (TN) = 12, False Positive (FP) = 12, False Negative (FN) = 7, and True Positive (TP) = 52. The high True Positive value indicates that the Random Forest model is highly effective in detecting children with stunting. Additionally, the relatively low False Negative value (7) suggests that only a small number of stunting cases go undetected by the model, indicating that the model is sufficiently sensitive to the stunting class.

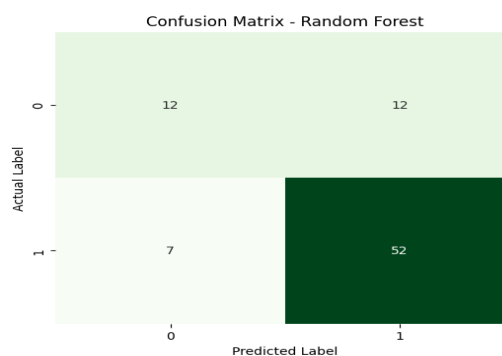


Figure 5. Random Forest Confusion Matrix Results

Based on the overall classification results, the Random Forest model achieved an accuracy rate of approximately 77%, which is higher than that of the previous XGBoost model. This indicates that the Random Forest model is more effective at distinguishing between stunted and non-stunted children in the dataset used. Additionally, this model demonstrates a good balance between its ability to detect the positive class (stunting) and minimize classification errors.

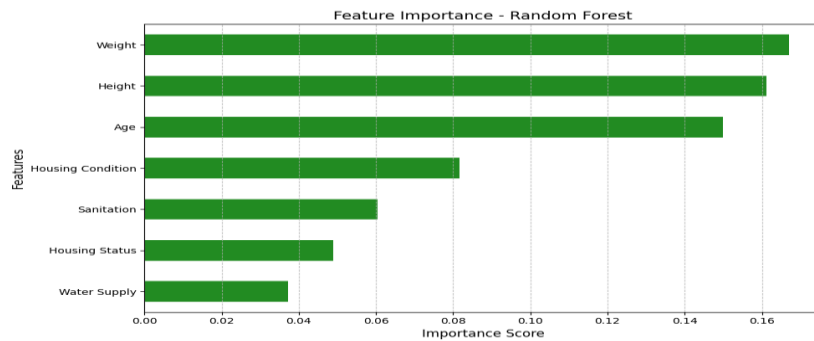


Figure 6. Random Forest Feature Importance Results

The feature importance results from the Random Forest model indicate that the most influential variables in stunting classification are dominated by the child’s physical factors—specifically, weight and height—which have the highest importance scores compared to other variables. This suggests that a child’s anthropometric status serves as the primary indicator for determining stunting status in the Random Forest model. Additionally, the age variable also makes a significant contribution, indicating that a child’s growth is heavily influenced by specific age stages.

3.4. Performance Evaluation

The results of the XGBoost model performance evaluation using the Receiver Operating Characteristic (ROC) curve indicate that the model has fairly good classification capabilities.

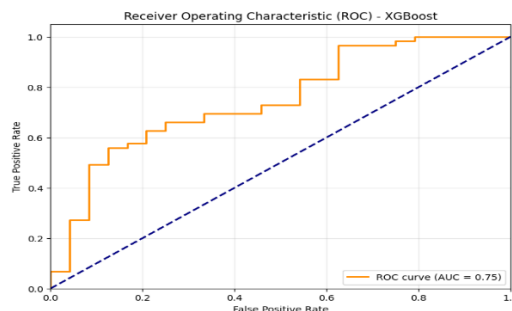


Figure 7. ROC and XGBoost Performance Evaluation Results

Based on the ROC plot, it can be seen that the curve lies above the diagonal line (random classifier), indicating that the model is better at distinguishing between the stunting and non-stunting classes than a random guess. The Area Under the Curve (AUC) value of 0.75 indicates that the model has a fairly good level of discrimination (good classification) in separating the two classes.

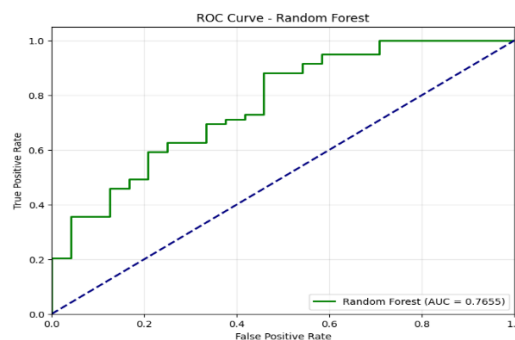


Figure 8. Performance Evaluation Results for ROC and Random Forest

Based on the ROC plot, the curve lies well above the diagonal line (baseline), indicating that the model has better predictive capabilities than random classification. The Area Under the Curve (AUC) value of 0.7655 indicates that the model has a good level of discriminative accuracy (good classification).

Table 1. Model Evaluation Results XGBoost

	Precision	Recall	F1-Score	Support
0	0.48	0.46	0.47	24
1	0.78	0.80	0.79	59
Accuracy			0.70	83
Macro Avg	0.63	0.63	0.63	83
Weighted avg	0.70	0.70	0.70	83

Table 2. Model Evaluation Results Random Forest

	Precision	Recall	F1-Score	Support
0	0.63	0.50	0.55	24
1	0.81	0.88	0.85	59
Accuracy			0.77	83
Macro Avg	0.72	0.69	0.70	83
Weighted avg	0.76	0.77	0.76	83

Based on the evaluation results in Tables 1 and 2, it can be seen that both models, XGBoost and Random Forest, performed reasonably well in classifying stunting, but at varying levels. The XGBoost model achieved an accuracy score of 0.70, with better performance in the stunting class (class 1) than in the non-stunting class (class 0). This is evident from the precision (0.78), recall (0.80), and F1-score (0.79) in the stunting class, indicating that the model is quite effective in detecting stunting cases. However, in the non-stunting class, the model's performance remained low, with an F1-score of only 0.47, indicating that the model still has difficulty accurately classifying non-stunting data.

Meanwhile, the Random Forest model performed better overall, with an accuracy score of 0.77. In the stunting class, this model has a precision value (0.81), recall (0.88), and F1-score (0.85), which are higher than XGBoost, thus indicating a better ability in detecting stunting cases. Furthermore, in the non-stunting class, Random Forest also experienced an increase in performance with an F1-score value of 0.55, although still considered moderate. The macro average value (0.70) and weighted average (0.76–0.77) in Random Forest were also higher than XGBoost, indicating that this model is more balanced in handling both classes.

4. DISCUSSIONS

The results of the study indicate that the application of the XGBoost and Random Forest algorithms in classifying stunting based on economic factors yields fairly good performance, with accuracy rates ranging from approximately 70% to 77%. Specifically, the Random Forest model demonstrates slightly better performance than XGBoost, both in terms of accuracy and AUC.AUC.

4.1. Performance and Data Complexity

The performance of the models in this study was influenced by the characteristics and complexity of the data used. Based on the evaluation results, the XGBoost and Random Forest models demonstrated fairly good performance with accuracy rates of approximately 70–77%, indicating that both algorithms were able to capture the patterns of relationships between economic variables and stunting status.

The complexity of the data is also evident from the imbalance in the amount of data between the stunting and non-stunting classes, which causes the models to tend to be more accurate in predicting the majority class.

4.2. The Role XGBoost dan Random Forest Optimasi

Optimization of the XGBoost and Random Forest algorithms plays a crucial role in improving the performance of stunting classification models. Essentially, both algorithms fall under ensemble learning methods, which combine multiple models to produce more accurate predictions. However, without optimization, the resulting model performance may not be optimal. In the XGBoost algorithm, optimization helps regulate the boosting process to more effectively minimize prediction errors. Meanwhile, in the Random Forest algorithm, optimization serves to improve the model's stability and generalization. Research findings demonstrate that optimization in both algorithms effectively enhances classification performance, including accuracy, precision, recall, and AUC. Additionally, optimization aids in identifying the features most influential on stunting. The differences in the characteristics of the two algorithms indicate that XGBoost is superior in improving accuracy through incremental learning, while Random Forest is superior in stability and resilience to data variations.data.

4.3. Limitations

This study has several limitations that should be considered when interpreting the results. One of the main limitations lies in the quantity and quality of the data used. Additionally, this study also faces the issue of an imbalanced dataset, where the number of stunting cases is greater than that of non-stunting cases. Another limitation concerns the variables used, as this study only considers a few basic economic and biological factors. In reality, stunting is a multidimensional issue also influenced by other factors such as nutritional intake, medical history, parenting practices, and access to healthcare services. In terms of methodology, this study used only two algorithms: XGBoost and Random Forest. Although both are robust algorithms, the use of other methods, such as deep learning or hybrid approaches, has not been explored. Furthermore, the parameter optimization process was limited, so it is possible that there are parameter combinations that could yield better performance.

5. CONCLUSION

Based on the results of the study, it can be concluded that classification models using the XGBoost and Random Forest algorithms are capable of identifying stunting status in children with fairly good performance. The evaluation results show that the models have an accuracy rate of approximately 70%, with Random Forest yielding superior results compared to XGBoost, particularly in terms of stability and more balanced classification performance. The higher AUC value of Random Forest also indicates better ability to distinguish between the stunting and non-stunting classes. From the results of the confusion matrix and classification report, it is known that the model is more optimal in detecting stunting cases compared to non-stunting cases, as indicated by the high recall value in the stunting class.

CONFLICT OF INTEREST

The authors declares that there is no conflict of interest between the authors or with research object in this paper.

ACKNOWLEDGEMENT

The authors would like to express their gratitude to the Politeknik Negeri Medan, which has consistently provided support for faculty facilities. The authors also thank the North Sumatra Health Office for providing the stunting dataset. Finally, we would like to thank the entire writing team for completing this paper.

REFERENCES

- [1] S. H. Marpaung, F. M. Sinaga, K. H. Rambe, and F. M. Sinaga, "Random Forest Optimization Using Recursive Feature Elimination for Stunting Classification," vol. 8, no. 1, pp. 281–287, 2025.
- [2] T. Sugihartono, B. Wijaya, A. F. Alkayes, and H. A. Anugrah, "Optimizing Stunting Detection through SMOTE and Machine Learning : a Comparative Study of XGBoost , Random Forest , SVM , and k-NN," vol. 6, no. 1, pp. 667–682, 2025.
- [3] A. Subekti, "Comparative Analysis of Automated Machine Learning Methods for Multiclass Stunting Prediction Using Anthropometric Data," vol. 10, no. 2, pp. 991–1002, 2026.
- [4] M. Anggito, H. Hadisuwarno, and K. T. Martono, "Komparasi performa model machine learning algoritma XGBoost dan Random Forest pada studi kasus mendeteksi stunting," vol. 22, no. 2, pp. 266–278, 2025.
- [5] D. M. Puspita, K. I. Santoso, A. Triyono, E. Supriyadi, and S. Agus, "Algoritma Random Forest , Decision Tree dan XGboost Untuk Klasifikasi Stunting Pada Balita," vol. 23, no. 1, pp. 67–76, 2025.
- [6] A. Birara *et al.*, "Prediction of stunting and its socioeconomic determinants among adolescent girls in Ethiopia using machine learning algorithms," pp. 1–27, 2025, doi: 10.1371/journal.pone.0316452.
- [7] M. I. Elim and E. Utami, "Performance Comparison of Child Stunting Prediction Support Vector Machine vs Random Forest with Grid Search Optimization," vol. 6, no. 5, pp. 5305–5319, 2025.
- [8] J. Pierre, G. Nduwayezu, P. Zhao, P. Pilesj, and A. Mansourian, "Environmental and Sustainability Indicators Multilevel small-area childhood stunting risk estimation : Insights from spatial ensemble learning , agro-ecological and environmentally remotely sensed indicators," vol. 27, no. July, 2025, doi: 10.1016/j.indic.2025.100822.
- [9] F. A. Wicaksana and C. Umam, "Comparative Analysis of Random Forest and Xgboost Performance for Network Flow Based Malware Classification," vol. 11, no. 1, pp. 108–115, 2026.
- [10] M. As and E. R. Subhiyakto, "Performance Comparison of Random Forest , SVM , and XGBoost Algorithms with SMOTE for Stunting Prediction," vol. 9, no. 4, pp. 1163–1169, 2025.
- [11] W. Richard, P. Thangata, B. Mkandawire, and N. Amoah, "Human Nutrition & Metabolism Advancing predictive analytics in child malnutrition : Machine , ensemble and deep learning models with balanced class distribution for early detection of stunting and wasting," *Hum. Nutr. Metab.*, vol. 42, no. June, p. 200340, 2025, doi: 10.1016/j.hnm.2025.200340.
- [12] M. Lailan, A. Azzawagama, A. Muslih, G. Zukhruf, and R. Sironpati, "Dialogues in Health Integrating education-based interventions and machine learning for stunting prevention : A case study in East Lombok , Indonesia," *Dialogues Heal.*, vol. 8, no. December 2024, p. 100264, 2026, doi: 10.1016/j.dialog.2025.100264.
- [13] F. Putrawansyah and M. Y. Idris, "Improving the Accuracy of Stunting Prediction in Children in Pagar Alam City Using XGBoost Feature Selection and K-Nearest Neighbor Classification," vol. 6, no. 6, pp. 5882–5898, 2025.
- [14] U. Children and P. New, "Machine Learning Algorithms for Predicting Stunting among Under-Five Children in Papua New Guinea," 2023.
- [15] H. G. Meles and J. Beyene, "Machine Learning Prediction of Child Stunting and Wasting in Ethiopia Using DHS Data : XGBoost and Random Forest Models with SHAP Interpretability."
- [16] S. Sallu, "Random Forest Machine Learning Analysis of Generative AI ' s Impact on Learning Effectiveness in Indonesian Higher Education," vol. 6, no. 5, pp. 3635–3651, 2025.
- [17] A. Hendy, R. K. Ibrahim, S. Mohammed, F. Abdelaliem, and A. Zaher, "Supervised machine

-
- learning for classification and prediction of stunting among under-five Egyptian children,” 2025.
- [18] I. Syahfitri, A. P. Juledi, and R. Muti, “Comparative Analysis of Machine Learning Algorithm Performance in Predicting Stunting in Toddlers,” vol. 8, no. 3, pp. 1452–1462, 2024.
- [19] N. Matovu, D. Ph, N. Kizito, M. Sc, M. Aljohani, and D. Ph, “Understanding household economic in fl uences on child stunting : A cross-sectional study in Eastern Uganda,” *Nutrition*, vol. 147, p. 113187, 2026, doi: 10.1016/j.nut.2026.113187.
- [20] M. K. Ayele, G. A. Baye, and S. H. Yesuf, “Predicting stunting status among under five children in ethiopia using ensemblemachine learning algorithms,” pp. 1–11, 2025.
- [21] T. Hariguna and A. Azis, “Health and Socio-Demographic Risk Factors of Childhood Stunting : Assessing the Role of Factor Interactions Through the Development of an AI Predictive Model,” vol. 5, no. 4, pp. 2175–2186, 2024.
- [22] A. R. Lubis, D. Wulandari, T. Ariyani, and Y. Y. Lase, “A COMPARATIVE STUDY OF PIPELINE-VALIDATED MACHINE LEARNING CLASSIFIERS FOR PERMISSION-BASED ANDROID MALWARE DETECTION,” vol. 20, no. 2, pp. 1675–1692, 2026.
- [23] A. Naik, “OPEN SGO enhanced random forest and extreme gradient boosting framework for heart disease prediction,” pp. 1–31, 2025.
- [24] Soni, M. A. Remli, K. M. Daud, and J. Al Amien, “Improving imbalanced class intrusion detection in IoT with ensemble learning and ADASYN-MLP approach,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 36, no. 2, pp. 1209–1217, 2024, doi: 10.11591/ijeecs.v36.i2.pp1209-1217.
- [25] M. S. Novelan and S. Aryza, “Optimization Cvrp With Machine Learning for Improved Classification of Imbalanced Data Food Distribution,” *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 10, no. 4, pp. 917–925, 2025, doi: 10.33480/jitk.v10i4.6467.